

Amendments to the Specification:

Please amend the paragraph beginning on page 11, line 7, as follows:

The raw document 71, in an arbitrary format, is converted 72 to a common internal tree-based representation. The representation may be described using an emerging standard called the DOM (Document Object Model) markup language, Document Object Model (DOM) Level 3 Core Specification, Version 1.0-<http://www.w3c.org>, but other formats are possible. For documents in some markup languages there exist publicly-available tools for performing this conversion, see The Tidy Project:

<http://www.w3.org/People/Raggett/tidy/>, but for documents in other markup languages, the conversion routine must be created *de novo*. Figure 13 shows an example of a simple HTML source document 88 and a corresponding tree-based representation (with the subtree underneath the table node omitted for clarity) 90.